Evolutionary Functional Genomics

1. Gene expression variation within and between populations

Population transcriptomics = the use of high-throughput methods that quantify gene expression levels (such as microarrays or RNA-seq) to determine levels of gene expression variation among individuals of a species. It is a combination of population genetics and transcriptomics.

Example:
Drosophila melanogaster has an ancestral species range in sub-Saharan Africa and has only recently spread to Europe and the rest of the world (within the past 15,000 years).

Questions:

a) How much expression variation is present in populations from the ancestral (African) and derived (European) species ranges?

b) Are there genes that show consistent expression differences between the populations? These genes may be important for adaptation to the new non-African environment.

It is important to note that gene expression levels are not constant within an individual. They may change over time (over development) or in response to the external environment. They also differ greatly between males and females. For this reason, comparisons are done using individuals of the same age, same sex, and raised under common conditions. Expression differences seen under such conditions are more likely to have a genetic basis.

2. Within-population expression polymorphism in males and females

In both male and female *D. melanogaster*, there are more gene expression differences between individuals from different populations than between individuals from the same population. On average, two European flies are more similar to each other in their gene expression than a European fly is to an African fly. Two African flies are more similar to each other in their gene expression than an African fly is to a European fly.

In males, there is less expression polymorphism on the X chromosome than on the autosomes. This pattern is not seen in females, where there is equal expression polymorphism on the X and autosomes. The reason for this appears to be related to male-biased gene expression: male-biased genes show the greatest expression polymorphism, but are significantly under-represented on the X chromosome. Since male-biased genes are mainly expressed in males and show very high expression polymorphism among males, their relative absence from the X leads to it having less expression polymorphism among males.

Overall, there is less expression polymorphism among females than among males. Several factors may contribute to this:

a) Genetic polymorphism on the Y chromosome has been shown to have a large effect on the expression of autosomal and X-linked genes. Because females lack a Y chromosome, this source of variation in absent in females.

b) Genetic polymorphism in the mitochondrial genome has been shown to have a large affect on male, but not female, gene expression. This may be because the mitochondria are inherited only from the mother and there is no direct selection against mutations that have a negative effect on male, but not female, gene expression.

c) It is possible that there is greater purifying selection to maintain an optimal expression level in females than in males.

3. Between-population expression polymorphism in males and females

When the African and European populations of *D. melanogaster* are compared, there are many more expression differences between populations in females than in males.

An early microarray study found 569 between-population expression differences in females, but only 153 in males. Furthermore, there was little overlap among the genes that showed between-population expression differences in males and females. Only 14 genes showed a significant between-population expression difference in both sexes. This suggests that gene expression evolves almost completely independently between males and females.

In both sexes, a cytochrome P450 gene involved in insecticide resistance (*Cyp6g1*) shows very high over-expression in European flies. This is caused by the insertion of a transposable element just upstream of the gene, as well as a tandem duplication of the gene.

Other genes that differ in expression between populations are involved in vision, olfaction, muscle formation, and behavior.

4. Functional analysis of putative *cis*-regulatory variants

If the expression divergence of a gene between African and European flies is the result of recent adaptive *cis*-regulatory evolution, then DNA sequence polymorphism in (or near) the gene should be affected in two ways:

a) DNA sequence polymorphism should be reduced in at least one of the populations, producing a signal of a "selective sweep"

b) There should be one or more fixed (or nearly fixed) sequence differences between the populations. That is, polymorphisms at high frequency in one population, but low frequency (or absent) in the other.

5. Example: *CG9509* (aka *fezzik*)

An example is the gene *CG9509*, which consistently shows 2–3 times higher expression in non-African ("cosmopolitan") flies than in African flies. There is an absence of sequence polymorphism in the genomic region just upstream of *CG9509* in the European population. There are also several fixed sequence difference between the European and the African populations in this region. These observations suggest that selection favored increased expression of *CG9509* in the European population and that the difference in expression is caused by sequence divergence in the *CG9509* upstream region.

How can this be tested?

A "reporter gene", in which the *CG9509* upstream region from either an African or a European allele is placed in front of the *E. coli lacZ* gene (encoding beta-galactosidase) can be constructed *in vitro*.

The reporter genes can then be inserted into a precise location of the *D. melanogaster* genome using a method known as PhiC31 site-specific integration. With this approach, the influence of the two different upstream regions (African and European) on gene expression can be compared in the same genetic background. If there is a difference in expression between the two reporter genes, then it must be the result of *cis*-regulatory divergence.

For *CG9509*, the European upstream region drives 2–3 times greater expression than the African upstream region. This indicates that the expression difference seen between the natural populations can be completely explained by *cis*-regulatory sequence divergence.

Interestingly, the SNP that has the greatest effect on *CG9509* expression is at intermediate frequency (40–50%) in cosmopolitan population, but at very low frequency (0–5%) in sub-Saharan African populations. This does not fit the classical model of a selective sweep, but

instead suggests that the polymorphism might be maintained by balancing selection in cosmopolitan populations.

The exact function of *CG9509* is unknown, but it is predicted to encode a choline dehydogenease enzyme and to be involved in ecdysteroid metabolism (growth hormone). It is expressed specifically in the Malphigian tubules, which are the analog of "kidneys" in insects. Studies using knock-out mutations or RNA interference have shown that *CG9509* expression influences stress tolerance and larval/adult growth. Because CG9509 are larger than wild-type flies, the gene has been named *fezzik*, after a giant character in *The Princess Bride*.

6. Example: *MtnA*

The brain transcriptomes have been compared between males and females and between European and African *D. melanogaster*. This revealed that there were more expression differences in the brain between populations than between sexes and most genes that differed in expression between populations showed the same pattern in both sexes. This is different from what is seen in whole flies.

A gene that is expressed differently between populations in the brain is the metallothionein gene *MtnA*, which is involved in heavy metal detoxification and oxidative stress tolerance. The expression difference appears to be caused by the deletion of a negative regulatory element in the *MtnA* 3' UTR, which may be a binding site for a microRNA. The deletion is in high frequency outside of Africa and shows evidence for positive selection in European populations. The deletion shows a clinal pattern on multiple continents, with the frequency increasing with distance from the equator. The deletion (and high *MtnA* expression) are associated with increased oxidative stress tolerance. As with *CG9509*, the deletion remains polymorphic in most cosmopolitan populations, suggesting that there may be a selective trade-off and it may be subject to balancing selection.

References:

Hutter, S., S. S. Saminadin-Peter, W. Stephan, and J. Parsch (2008) Gene expression variation in African and European populations of *Drosophila melanogaster*. Genome Biology 9: R12.

Müller, L., S. Hutter, R. Stamboliyska, S. S. Saminadin-Peter, W. Stephan, and J. Parsch (2011) Population transcriptomics of *Drosophila melanogaster* females. BMC Genomics 12: 81.

Parsch, J. (2011) The cost of being male. Science 332: 798-799.

Catalán, A., S. Hutter, and J. Parsch (2012) Population and sex differences in *Drosophila melanogaster* brain gene expression. BMC Genomics 13: 654.

Catalán, A., A. Glaser-Schmitt, E. Argyridou, P. Duchen, and J. Parsch (2016) An indel polymorphism in the *MtnA* 3' untranslated region is associated with gene expression variation and local adaptation in *Drosophila melanogaster*. PLoS Genet. 12(4):e1005987.

Glaser-Schmitt A., and J. Parsch (2018) Functional characterization of adaptive variation within a *cis*-regulatory element influencing *Drosophila melanogaster* growth. PLoS Biol 16: e2004538.